アルゴリズム 60－1
（1998．1．22）

連結グラフのランダム生成法について

矢農 正紀　　　二村 良彦
早稲田大学大学院理工学研究科　早稲田大学理工学部情報学科
〒169–0072 東京都新宿区大久保3–4–1
e-mail：{yanoh,futamura}@futamura.info.waseda.ac.jp

概要

グラフアルゴリズムの精密な評価を行うためには、指定された頂点数nと辺数mを有し、かつランダムに生成された複数のラベル付き連結グラフが必要である。本稿では、頂点にラベルが付くか、辺に異なる重みが付く場合に、頂点n個かつ辺m本の連結グラフを一様ランダムに生成するアルゴリズムを提案する。このアルゴリズムは、構成比と呼ばれる部分グラフ数のグラフ数に対する割合を、計算機オーバーフローを回避して直接計算する方法に基づく。ランダム生成の前処理として、表の作成に$O(m^2n^2)$時間と$O(mn^2)$空間を要するが、これにより連結グラフを1個につき$O(mn)$時間で生成することが可能である。

Random Generation of Connected Graphs

Masanori Yanoh and Yoshihiko Futamura
Graduate School of Science and Engineering, Waseda University
Department of Computer and Information Science,
Waseda University
3–4–1 Okubo Shinjuku-ku, Tokyo 169–0072 Japan
e-mail：{yanoh,futamura}@futamura.info.waseda.ac.jp

Abstract

This paper proposes a new method for generating random labeled connected graphs with n nodes and m edges for given n and m. The random graphs are to be used to conduct precise evaluation of graph algorithms. Our method is based on overflow free calculation of a composite ratio, i.e. the ratio of number of subgraphs against the number of all graphs. Using the ratio, we can uniformly generate a connected graph with n nodes and m edges in $O(mn)$ time using preprocessing of $O(m^2n^2)$ time $O(mn^2)$ space.

1 はじめに

アルゴリズムの教育及び研究に際して、アルゴリズムが扱う各種データをランダムに供給するランダムデータサーバー (RDS) が必要である [1]。我々は既に、ソーティングアルゴリズム等の評価のためにある種の性質を有する順列を生成するRDSを開発し、教育と研究に利用している [2][3][4]。本稿は、各種グラフをランダムに生成するRDSを実現するための理論的基礎に関するものである。

最小木問題や最短経路問題を解くグラフアルゴリズムの評価を精密に行うためには、指定された頂点数nと辺数mを有し、かつランダムに生成された複数のラベル付き連結グラフが必要である。本稿で扱うグラフは、多重辺やループを含まない単純な無向グラフである。

頂点にラベルが付いた連結グラフについては、頂点数nのみを指定する生成アルゴリズムは知られている [9]。しかし、生成したグラフをテストデータとして利用するためには辺数mも指定する必要がある。そのためには、n個の頂点をランダムにm本の辺で結び、それが連結グラフとなるまで同様の操作を繰り返す方法が考えられる。しかしながら、この方法は辺数が小さい場合には、時間が掛かり過ぎて実用的ではない。また、頂点数nの全域木を生成して、辺数がmになるまでランダムに辺を追加していく方法も考えられる。しかし、この方法では生成されるグラフの一様性が保証されない。この2点については、本稿において詳しく議論する。

本稿では、連結グラフの総数を求める再帰方程式を用いて、頂点数n辺数mの連結グラフを一様性を保証してランダムに生成するアルゴリズムを提案する。グラフの総数に対して、特定の性質を持つ部分グラフが占める割合を構成比と呼ぶ。構成比は、ランダム生成を行う上での確率に対応し、再帰的な生成を行うためにその値が必要となる。本稿では、最初に連結グラフの総数を求める再帰方程式 [14] について説明する。次に、その再帰方程式に基づいて構成比を定義し、かつ構成比を計算機オーバーフローを起こさずに計算する構成比直接計算法を示す。その上で、構成比を用いた$O(m^2n^2)$時間と$O(mn^2)$空間の前処理法及び$O(mn)$時間の連結グラフ生成アルゴリズムについて述べる。

さらに、ラベル付き連結グラフの生成法を利用して、重み付き連結グラフの一様生成が直ちに可能となることを示す。ラベル無しグラフのランダム生成 [13] に関する研究は既に行われている。しかし、多くのグラフアルゴリズムの精密な評価を行うためには、頂点にはラベルが無いが各辺に異なる重みが付いた連結グラフの一様生成が必要である。

最後に、生成した連結グラフに対して一意な番号を付加する関数を定義する。そして、その関数を利用してχ^2検定を行う方法とその結果についても報告する。これにより、ランダムに生成された連結グラフに対して、その一様性を検証することが可能となる。

2 ラベル付き連結グラフの再帰方程式

頂点n個かつ辺m本の単純無向グラフで、各頂点に対し$v_1, v_2, \ldots, v_n$とラベルが付いた連結グラフの総数を$G(n,m)$とする。$G(1,0) = 1$であり、$m < n-1$または$m > \binom{n}{2}$の場合は$G(n,m) = 0$である。

$m = n-1$の場合、グラフは木となるので$G(n, n-1) = n^{n-2}$が成り立つ [8]。$m = n$については、Rényi [11] によって、以下の式が求められている。ここでrは、グラフにただ一つ存在する閉路の長さに対応している。

$$G(n,n) = \sum_{r=3}^{n} \binom{n}{r} \frac{(r-1)!}{2} r n^{n-r-1} \tag{1}$$

一般の$G(n,m)$については、Wormald[14] によって、以下の再帰方程式が求められている。

$$mG(n,m) = \left[\binom{n}{2} - (m-1)\right] G(n, m-1) + \frac{1}{2} \sum_{i=1}^{n-1} \sum_{j=0}^{m-1} F(n,m,i,j) \tag{2}$$

$$F(n,m,i,j) = \binom{n}{i} i(n-i) G(i,j) G(n-i, m-j-1)$$

式(2)の左辺は、1本の辺に印を付けた連結グラフの総数である。右辺の第1項は、印を付けた辺を削除してもなお連結しているグラフの総数を表す。そして、第2項の二重和は、印を付

けた辺が2つの連結グラフの橋となるグラフの総数を表す[10]。

3 構成比直接計算法

本節では、連結グラフの生成を行う上で必要となる構成比を、(n,n)-連結グラフと一般の連結グラフについて定義し、構成比を直接計算する方法を示す[5][6]。直接計算を行うことで、$G(n,m)$ の値を求めることにより発生する計算機オーバーフローを回避できる。

構成比の直接計算を行うため、$n \le m \le \binom{n}{2}$ について $R(n,m)$ を以下の式で定義する。

$$R(n,m) = \frac{G(n,m)}{G(n,m-1)}$$

$m < n$ または $m > \binom{n}{2}$ の場合は $R(n,m) = 0$ とする。$R(n,m)$ を求めることにより、対応する構成比の値を得ることができる。

3.1 (n, n)-グラフの構成比直接計算法

式(1)から、頂点 n 個かつ辺 n 本のラベル付き連結グラフに対する構成比 $Q(n,r)$ を、$r \ge 3$ について以下の式で定義する。

$$Q(n,r) = \frac{\binom{n}{r} r! n^{n-k-1}}{2G(n,n)}$$

この構成比 $Q(n,r)$ は、(n,n)-連結グラフが長さ r の閉路をもつ確率に対応する。$Q(n,r)$ を求めるため、$R(n,n)$ の評価を行うと以下の結果を得る。

$$R(n,n) = \frac{G(n,n)}{G(n,n-1)} = \sum_{r=3}^{n} J(n,r) \qquad (3)$$

$$J(n,r) = \frac{\binom{n}{r} r! n^{n-r-1}}{2n^{n-2}} = \frac{n}{2}\prod_{k=1}^{r-1}\left(1-\frac{k}{n}\right) \quad (4)$$

式(3)(4)を用いて $J(n,r)$ と $R(n,n)$ を計算することで、以下の式で構成比 $Q(n,r)$ の値を得ることができる。

$$Q(n,r) = \frac{J(n,r)}{R(n,n)}$$

$Q(n,r)$ の値を得るために必要となる計算量は、各々の n について $O(n^2)$ 時間である。

3.2 一般グラフの構成比直接計算法

頂点 n 個かつ頂点 m 本のラベル付き連結グラフの構成比 $P(n,m,i,j)$ を、式(2)から以下のように定義する。

$$P(n,m,0,0) = \frac{\left[\binom{n}{2} - (m-1)\right] G(n,m-1)}{mG(n,m)}$$

$$P(n,m,i,j) = \frac{F(n,m,i,j)}{2mG(n,m)}$$

この構成比は、連結グラフに対し特定の部分構造が存在する確率に対応する。さらに、式(2)を用いて $R(n,m)$ の式を評価する。

$$R(n,m) = \frac{\binom{n}{2} - (m-1)}{m} + \sum_{i=1}^{n-1} \sum_{j=f(n,m,i)}^{g(n,m,i)} K(n,m,i,j)$$

$$K(n,m,i,j) = \frac{1}{2m}\frac{n^{\underline{i+1}}}{(i-1)!}\frac{G(i,j)G(n-i,m-j-1)}{G(n,m-1)}$$

ここで、$n^{\underline{k}} = n(n-1)\cdots(n-k+1)$ と定義する。また、$f(n,m,i) = \max(i-1, m-1-\binom{n-i}{2})$ かつ $g(n,m,i) = \min(\binom{i}{2}, m-n+i)$ である。

$R(n,m)$ を計算するため、$K(n,m,i,j)$ の値を求める必要がある。以下の各式に従うことにより、$G(n,m)$ の値を用いることなく $K(n,m,i,j)$ を直接計算することができる[5]。

1. $m = n$ の場合
   (a) $j = i-1$ の場合

$$K(n,n,i,i-1) = \left[\prod_{r=1}^{i-1}\frac{i}{r+1}\left(1-\frac{r}{n}\right)\right]\left(1-\frac{i}{n}\right)^{n-i-1} \cdot \frac{R(n-i,n-i)}{2} \qquad (5)$$

   (b) $j = i$ の場合

$$K(n,n,i,i) = \left[\prod_{r=1}^{i-1}\frac{i}{r+1}\left(1-\frac{r}{n}\right)\right]\left(1-\frac{i}{n}\right)^{n-i-1} \cdot \frac{R(i,i)}{2} \qquad (6)$$

2. $m > n,\ n/2 < i < n,\ m-j = n-i$ の場合

$$K(n,m,i,j) = K(n,m,n-i,n-i-1) \qquad (7)$$

3. $m > n,\ j = f(n,m,i)$ の場合

(a) $f(n,m,i) = m - 1 - \binom{n-i}{2}$ の場合

$$
\begin{aligned}
&K(n,m,i,j)\\
&= \frac{m-1}{m(m-j-1)R(n,m-1)}\\
&\quad \cdot K(n,m-1,i,j) \qquad (8)
\end{aligned}
$$

(b) $f(n,m,i) = i - 1$ の場合

$$
\begin{aligned}
&K(n,m,i,i-1)\\
&= \frac{(m-1)R(n-i,m-i)}{mR(n,m-1)}\\
&\quad \cdot K(n,m-1,i,i-1) \qquad (9)
\end{aligned}
$$

4. $m > n,\ f(n,m,i) < j \leq g(n,m,i)$ の場合

$$
\begin{aligned}
&K(n,m,i,j)\\
&= \frac{R(i,j)}{R(n-i,m-j)}\\
&\quad \cdot K(n,m,i,j-1) \qquad (10)
\end{aligned}
$$

式(5)(6)により求まる$K(n,n,i,j)$の値を用いることで、式(7)(8)(9)(10)に従って値の小さいnとmから順に$K(n,m,i,j)$と$R(n,m)$を相互に計算して値を得ることが可能となる。

$R(n,m)$を求めるために必要となる計算量は、$G(n,m)$の計算と同様に$O(m^2n^2)$時間となる。最低限必要となる空間に関しては、$R(n,m)$と$K(n,m,i,f(n,m,i))$について値の表が必要となるので、合計して$O(mn^2)$空間となる。

$K(n,m,i,j)$と$R(n,m)$の値を用いることで、以下の式により構成比$P(n,m,i,j)$を計算することができる。

$$
P(n,m,0,0) = \frac{\binom{n}{2} - (m-1)}{mR(n,m)}
$$
$$
P(n,m,i,j) = \frac{K(n,m,i,j)}{R(n,m)}
$$

4 構成比を用いた生成アルゴリズム

前節で示した構成比は、任意の連結グラフに対して特定の性質を持つ部分グラフが存在する確率に対応するので、直接計算法に従って求めた構成比の値を用いて、頂点n個かつ辺m本のラベル付き連結グラフを等確率で、すなわち各々の連結グラフを確率$1/G(n,m)$で生成することが可能となる。

表 1: 正規表現と変換の例

x_i	4	4	3	7	4	7
v_i	1	2	5	3	6	4

表 2: 拡張正規表現と変換の例

x_i	2	5	7	2	6	5	2
v_i	1	3	4	7	2	6	5

本節では、頂点n個かつ辺m本のラベル付き連結グラフの生成法を、$m = n-1$と$m = n$の場合について述べ、それから一般の$m > n$について生成アルゴリズムを示す。

4.1 (n, n − 1)-連結グラフの生成

$m = n - 1$の場合、連結グラフは木となるので、正規表現[8]の変換を行うことで、一様生成が可能となる。自由度n、長さ$n-2$の乱数列$x_1, x_2, \ldots, x_{n-2}$を一様生成して、$x_{n-1} = n$とする。このとき、$x_1, \ldots, x_{n-2}, x_{n-1}$を木の正規表現と呼ぶ。以下で、この正規表現の木への変換手順[8]を示す。

まず、この数列に含まれない最小の番号をv_1とする。そして、$1 < i \leq n$について$v_1, \ldots, v_{i-1}$と$x_i, \ldots, x_{n-1}$に含まれていない最小の番号をv_iとすると、長さ$n-1$の順列$v_1, v_2, \ldots, v_{n-1}$が得られる。最後に、辺$(v_i, x_i)$を接続させることで木が得られる。表1に頂点数$n = 7$の場合の正規表現とその変換の例を示す。この変換は$O(n^2)$で容易に可能である。

正規表現の総数はn^{n-2}通りあり、各々の正規表現と木が一対一対応するので、$(n, n-1)$-連結グラフが一様に生成される。

4.2 (n, n)-連結グラフの生成

$m = n$の場合、グラフは閉路を1つだけもつ。以下の生成アルゴリズムは、まず構成比を用いて閉路の長さを決定し、それから前述した正規表現を拡張して連結グラフを生成する。

実行の効率化を図るため、構成比の和を次のように定義する。ただし、$r \geq 3$とする。

$$
U(n,r) = \sum_{k=3}^{r} Q(n,k)
$$
$$
U(n,2) = 0
$$

この構成比の和$U(n,r)$を用いて、$p \in [0,1)$の値をとる一様乱数pを生成して、対応するr、すなわち$U(n,r-1) \le p < U(n,r)$を満たすrを二分探索法[12]を用いて$O(\log n)$で決定することができる。このとき、各々のrは確率$Q(n,r)$で選択されているので、引き続いて閉路長rの(n,n)-連結グラフを生成すればよい。

閉路長rの(n,n)-連結グラフの生成は、拡張した正規表現を変換することで行う。まず、自由度n、長さ$n-r-1$の乱数列$x_1, x_2, \ldots, x_{n-r-1}$を生成する。次に、$\{1,2,\ldots,n\}$から$r$個の要素を選び、それらをランダムに並び変えたものを$x_{n-r}, \ldots, x_{n-1}$とする。このとき、$x_{n-r} > x_{n-1}$ならば逆順に並び変え$x_{n-r} < x_{n-1}$が成り立つようにする。最後に、$x_n = x_{n-r}$とする。こうして得られた$x_1, x_2, \ldots, x_n$を拡張正規表現と呼ぶことにする。以下で、拡張正規表現を(n,n)-連結グラフへ変換する手順を示す。

$m = n-1$の場合と同様に、数列に含まれない最小の番号をv_1、$1 < i \le n-r$について$v_1, \ldots, v_{i-1}$と$x_i, \ldots, x_n$に含まれていない最小の番号をv_iとして$v_1, v_2, \ldots, v_{n-r}$を得る。$n-r < i \le n$については、$v_i = x_{i-1}$とすることで、長さ$n$の順列$v_1, v_2, \ldots, v_n$が得られ、辺$(v_i, x_i)$を接続させて連結グラフを生成する。表2に頂点数$n=7$、閉路長$r=3$の場合の拡張正規表現とその変換の例を示す。$m=n-1$の場合と同様に$O(n^2)$時間で生成が可能である。

4.3 一般連結グラフの生成

$m > n$の連結グラフについても、構成比の和を定義する。まず、部分グラフの頂点数iに対応する構成比を以下の式で定義する。

$$S_1(n,m,0) = P(n,m,0,0)$$
$$S_1(n,m,i) = S_1(n,m,i-1) + \sum_{j=f(n,m,i)}^{g(n,m,i)} P(n,m,i,j)$$

次に、部分グラフの辺数jに対応する構成比を以下の式で定義する。

$$S_2(n,m,i,j) = \sum_{k=f(n,m,i)}^{j} P(n,m,i,k)$$

これらの構成比の和の値と構成比の値を用いることで、連結グラフの一様生成を行うことができる。アルゴリズムを以下に示す。

(構成比生成アルゴリズム)

1. $n < 2$ならば終了。さもなければ$k := m$として、一様乱数$p \in [0,1)$を発生させる。
2. もし$p \ge S_1(n,k,0)$であれば3.に進む。さもなければ、kの値を1減らして繰り返す。
3. $k \le n$であれば正規表現を用いてグラフを生成して終了。さもなければ4.に進む。
4. $S_1(n,k,i-1) \le p < S_1(n,k,i)$となる$i$の値を二分探索法を用いて表より決定する。
5. n,k,iを固定して、jについて$P(n,k,i,j)$の値と和$S_2(n,k,i,j)$を計算して表を作成し、$S_2(n,k,i,j-1) \le p < S_2(n,k,i,j)$となる$j$の値を決定する。
6. ランダムにi個の頂点を選んで、頂点i個かつ辺j本の連結グラフを再帰的に生成する。
7. 他の頂点について、頂点$n-i$個かつ辺$k-j-1$本の連結グラフを再帰的に生成する。
8. 6.と7.で生成した2つの連結グラフから、それぞれ頂点v, wを1個ずつランダムに選び、辺(v,w)で連結させる。
9. 8.で生成した連結グラフに対し、接続していない辺から$m-k$本の辺を選び、接続させる。

このアルゴリズムに必要な空間は、構成比の和$S_1(n,m,i)$と構成比の表の大きさに従う。構成比の値は4.の計算部分で必要となるが、$R(n,m)$と$K(n,m,i,f(n,m,i))$の表があれば、前述の直接計算法により構成比$P(n,k,i,j)$と和$S_2(n,k,i,j)$をアルゴリズム実行中に$O(m)$時間で求めることができる。以上から、$O(mn^2)$空間を必要とすることがわかる。実行時間に関する結果を以下に示す。

(定理1) 構成比生成アルゴリズムの実行時間は$O(mn)$である。

(証明) 頂点n個かつ辺がm本以下で、最も生成に時間を要する連結グラフに対する、9.の部分を除いた実行時間を$T(n)$とする。再帰的に生成を行う場合、非再帰部分で最も時間を要するのは5.の計算部分で、$O(m)$時間であるから、以下の再帰方程式が成り立つ。

$$T(n) \le cm + T(k) + T(n-k)$$

ここで、cは定数であり、$1 < k < n$である。従って、9. を除いた部分の実行時間は$O(mn)$で抑えられる。

そして、9. の部分については、グラフの各頂点の次数をデータとして保持することで、辺1本につき$O(n)$時間で実行可能である。付加する辺の数は高々m本であるから、全体を通しての9. の実行時間も$O(mn)$で抑えられる。□

5 重み付き連結グラフのランダム生成

本節では、辺に異なる重み$w_1, w_2, \ldots, w_m$が付けられ、頂点にはラベルが付いていない重み付きグラフを考える。頂点n個かつ辺m本の重み付き連結グラフの総数を$H(n,m)$とすると、$n \geq 3$について$m!G(n,m) = n!H(n,m)$が成り立つ [7]。これより、以下の定理が成立する。

(定理2) 任意のラベル付き連結グラフが確率$1/G(n,m)$で生成されれば、そのグラフの各辺に一様ランダムな重み付けを行って頂点を消し去ることで、重み付きグラフを一様ランダムに生成できる。

(証明) 重み付けが$m!$通りあるので、生成され得るグラフの組合せは$m!G(n,m)$通りとなる。そして、頂点を消し去ることで同一となるグラフが$n!$通りあるので、任意の重み付きグラフが生成される確率は以下の式で与えられる。

$$\frac{n!}{m!G(n,m)} = \frac{1}{H(n,m)}$$

従って、重み付き連結グラフが$1/H(n,m)$の等確率で、一様に生成されることがわかる。□

以上により、重み付き連結グラフの一様生成は、前述したアルゴリズムを用いてラベル付き連結グラフを生成してから、定理2の手順に従うことで直ちに可能となる。

6 ラベル付き連結グラフの各種生成法

本節では、構成比生成アルゴリズム以外に考えられる、ラベル付き連結グラフのランダム生成に関して述べる。

まず、以下のように生成と検査を繰り返す素朴なアルゴリズムが考えられる。

(生成検査法)

1. $\binom{n}{2}$本の辺から、ランダムにm本の辺を選んでグラフを生成する。
2. 生成したグラフに対し深さ優先探索を行う。
3. 全ての頂点が連結していれば終了。さもなければ1. に戻って再度繰り返す。

このアルゴリズムでは、連結グラフとなるまで生成を繰り返し、その回数は負の二項分布に従う。また、隣接行列とリスト表現を組合せれば1回の生成が平均$O(m)$で可能なので、1個の連結グラフの生成に必要な実行時間は平均$O(n^2 + m/p_{n,m})$とできる。ここで、$p_{n,m}$は頂点n個かつ辺m本のグラフが連結している確率であり、以下の式で表せる。

$$p_{n,m} = G(n,m)/\binom{\binom{n}{2}}{m}$$

生成検査法の問題点としては、mがnに近い値で、グラフの連結している確率が低い場合に実行時間が増大することがあげられる。例えば、頂点50個、辺52本の場合、$p_{50,52} \approx 2.19 \cdot 10^{-5}$であるから、平均して約45600回もの生成を繰り返すことになる。

また、以下のアルゴリズムは、連結グラフを一見公平に生成するように思える。

(全域木法)

1. 頂点n個の木をランダムに生成する。
2. 木に$m - (n-1)$本の辺を付加する。

木を生成する部分は、前述した正規表現の変換を行えばよい。辺を付加する部分と合わせて、実行時間は$O(n^2)$とできる。

この全域木法は、全ての連結グラフを生成することができるが、確率的に偏った生成を行うため、簡便法といえる。例えば、頂点5個、辺6本の場合、簡便法では$5^3 \cdot \binom{6}{2} = 1875$通りの生成がなされるが、実際には$G(5,6) = 205$で$1805 \bmod 205 \neq 0$なので、各々の連結グラフの生成される確率が一様とはならない。

全域木法の系として、(n,n)-連結グラフの構成比を用いた以下のアルゴリズムも考えられる。

((n, n)-法)

1. 頂点n個、辺n本の連結グラフをランダムに生成する。
2. 生成したグラフに$m - n$本の辺を付加する。

一般グラフの場合と異なり、構成比の計算に前処理を必要とせずに$O(n^2)$時間で生成可能であるが、一様性が保証されないのは全域木法と同様である。

7 連結グラフに対するχ^2検定

ラベル付き連結グラフが一様に生成されたか否かを調べるためには、χ^2検定を用いることができる。すなわち、連結グラフに1から$G(n,m)$までの番号を付け、それらの列に対してχ^2検定を行えばよい。本節では、連結グラフの番号付けの方法に関して述べ、それから検定の結果を報告する。

7.1 連結グラフの番号付け

ラベル付き連結グラフの番号付けには、式(2)とは別の再帰方程式を用いる。以下の$G(n,m)$を計算する再帰方程式[7]は、文献[9]で与えられている式に対して、辺の数も指定できるよう一般化を行ったものである。

$$G(n,m)=\sum_{k=0}^{n-1}\sum_{j=1}^{k}\sum_{i=k-1}^{m-j}Y(n,m,i,j,k) \quad (11)$$

$$Y(n,m,i,j,k)=\binom{n-2}{k-1}\binom{k}{j}G(k,i)G(n-k,m-i-j)$$

(証明) 頂点n個かつ辺m本の連結グラフ$G_{n,m}$の2頂点v_1とv_nに注目し、頂点v_nから出る辺を全て$G_{n,m}$から取り去ったとき、頂点v_1と連結している頂点の数をkとする。これらの頂点からなる部分グラフの辺の数をi、部分グラフを$H_{k,i}$とする。頂点の選び方から$H_{k,i}$は連結グラフとなるから、その総数は$G(k,i)$であり、頂点の組合せが$\binom{n-2}{k-1}$通りある。

次に、頂点v_nと$H_{k,i}$の頂点間の辺数をjとすると、$1\le j\le k$が成立し、その組合せは$\binom{k}{j}$通りとなる。

そして、$H_{k,i}$に含まれない$n-k$個の頂点のうち、v_n以外の頂点は$H_{k,i}$の頂点と接続していない。さらに、$n-k$個の頂点と$m-i-j$本の辺からなる部分グラフ$H_{n-k,m-i-j}$は連結グラフなので、その総数は$G(n-k,m-i-j)$である。

よって、k,i,j各々について、以上の和をとったものが連結グラフの総数となる。□

また、$\{1,2,\dots,n\}$から$C=\{s_1,s_2,\dots,s_r\}$を選ぶ組合せに対し、以下の関数i_cを用いて0から$\binom{n}{r}-1$の番号を付けることができる。ここで、$s_k<s_{k+1}$ $(1\le k<r)$で昇順と仮定する。

$$i_c(C)=\sum_{k=1}^{r}\binom{s_k-1}{k}$$

番号付けを行うため、生成した連結グラフGに対して、前述の証明での手順と同様にi,j,kと部分グラフ$H_{k,i},H_{n-k,m-i-j}$を求め、v_1,v_nを除いた$n-2$個の頂点のうち、$H_{k,i}$の$k-1$個の頂点の組合せをC_1、$H_{k,i}$の頂点を$1,2,\dots,k$とつけかえ、v_nと接続しているj本の頂点の組合せをC_2とする。

各(i,j,k)の値に対応する連結グラフの個数は$Y(n,m,i,j,k)$であるから、以下で定義する関数iにより番号付けを行うことができる。

$$\begin{aligned}
i(G_{1,0}) &= 1\\
i(G_{n,m}) &= \sum_{r=0}^{k-1}\sum_{q=1}^{r}\sum_{p=r-1}^{m-q}Y(n,m,p,q,r)\\
&\quad+\sum_{q=1}^{j-1}\sum_{p=k-1}^{m-q}Y(n,m,p,q,k)\\
&\quad+\sum_{p=k-1}^{i-1}Y(n,m,p,j,k)\\
&\quad+\Bigg(\bigg(i_c(C_1)\binom{k}{j}+i_c(C_2)\bigg)G(k,i)\\
&\qquad+i(H_{k,i})-1\Bigg)G(n-k,m-i-j)\\
&\quad+i(H_{n-k,m-i-j})
\end{aligned}$$

7.2 検定の結果

前項で求めた番号付け関数iを用いて、構成比生成アルゴリズム、生成検査法、全域木法と(n,n)-法の4方式により、頂点数7かつ辺数8のラベル付き連結グラフ$(G(7,8)=156555)$を各5000000個生成し、各々のグラフより番号を求めて作成した数列に対してχ^2検定を行った。その結果および各方式の計算量を表3に示した。

文献[12]では、数列を乱数列とみなせる範囲は、統計量が156555から約±791以内に収まることとする目安を与えている。表3において、理論的に一様性が保証されている構成比生成アルゴリズムと生成検査法は、共に良い結果を示している。一方、全域木法は、統計量が有意に範囲の外にあるため、連結グラフを一様に生成するには不適といえる。(n,n)-法に関しても、全

表 3: 頂点7個、辺8本のラベル付き連結グラフに対して行ったχ^2検定の結果

方式	理論的実行時間	χ^2統計量
構成比	最悪$O(mn)$	157069.5
生成検査	平均$O(n^2 + m/p_{n,m})$	156227.8
全域木法	最悪$O(n^2)$	509530.7
(n, n)-法	最悪$O(n^2)$	255784.4

域木法よりは小さい統計量を得ているが、全域木法と同様有意に範囲外であり、一様に生成するには不適といえる。

頂点数がより多い場合には、例えば頂点9個、辺15本の場合$G(9, 15) = 5228627544$となるように、連結グラフの総数が大きな数となるので、χ^2検定で一様性を検証することは難しくなる。また、このような大きな自由度を持つ乱数列の検定法は、我々の調査の範囲では見出すことが出来なかった。

構成比生成アルゴリズムの問題点は、前処理に必要な時間が$O(m^2n^2)$と大きいことである。また、生成検査法では辺数が小さい場合、すなわちmがnに近い場合に実行時間が増大する。一方、全域木法では実行時間は増大しないが、生成された連結グラフの一様性が保証されない。従って、mがnに近い場合に、$O(n^2)$時間での一様生成が可能な、構成比を利用した生成アルゴリズムは有効であると考えられる。

8 おわりに

本稿では、頂点数nと辺数mを指定した連結グラフのランダム生成を構成比を用いて行う方法を提案し、かつその有効性について検討した。我々の提案した方法は、辺数mが頂点数nと同じオーダーのとき、他方式より高性能な連結グラフ生成法であることを示した。

本稿で提案した連結グラフのランダム生成法に基づくRDSの実用化(即ち、高速化)が今後の課題である。そのためには、構成比に関する精度の高い近似式を発見する必要がある。また、有向グラフを含めた、他のクラスのグラフに対するランダム生成アルゴリズムを開発することも今後の課題である。

参考文献

[1] 二村, 根本, 矢農, アルゴリズム教育のためのランダムデータサーバー, 第11回私情協大会事例集, E-3(1997).

[2] 二村, 大谷, 青木, 二村, 単純指標を持つ乱順列の高速生成法, コンピュータソフトウェア, Vol. 14, No. 6(1997).

[3] 二村, 高野, 青木, 特性指標を制御した乱順列生成方法およびそのプログラム, 特願平8-250030(1996).

[4] 二村, 高野, 青木, 矢農, ランダムデータジェネレータ生成方式およびそのプログラム, 特願平9-550240(1997).

[5] 二村, 高野, 矢農, 連結グラフの生成方法およびそのプログラム, 特許平9-259958(1997).

[6] 二村, 矢農, 実際的プログラム変換の例, 日本ソフトウェア科学会第14回大会論文集, E8-2(1997).

[7] 矢農, 二村, 連結グラフのランダム生成, 日本ソフトウェア科学会第14回大会論文集, B12-3(1997).

[8] D.E.Knuth, *The Art of Computer Programming*, Vol.1, Addison-Wesley, Reading, Massachusetts(1973).

[9] A.Nijenhuis and H.S.Wilf, The Enumeration of Connected Graphs and Linked Diagrams, *J. Combin. Theory Ser. A* 27(1979) 356-359.

[10] E.M.Palmer, *Graphical Evolution – An Introduction to the Theory of Random Graphs*, Wiley, New York(1985).

[11] A.Rényi, On Connected Graphs I, *Publ. Math. Inst. Hungarian Acad. Sci.* 4(1959) 385-388.

[12] R.Sedgewick, *Algorithms in C*, Addison Wesley, Reading, Massachusetts(1990).

[13] N.C.Wormald, Generating Random Unlabelled Graphs, *SIAM J.Comput.*, Vol. 16, No. 4(1987) 717-727.

[14] N.C.Wormald, *Some Problems in the Enumeration of Labelled Graphs*, Ph.D. thesis, University of Newcastle(1978).